

基于机器学习的商业银行绿色信贷风险管理研究

汪 婷, 张 莉

扬州大学商学院, 江苏 扬州

收稿日期: 2025年3月12日; 录用日期: 2025年3月27日; 发布日期: 2025年4月27日

摘 要

本文通过构建基于大数据的综合评价体系, 探讨提升绿色信贷风险评估效率和准确性的途径。选取了我国470家上市公司作为样本, 运用SelectFromModel模型筛选关键指标, 并结合SMOTE过采样算法处理样本不均衡问题。研究结果得出随机森林算法采用的机器学习模型更适用于评估绿色信贷风险。

关键词

绿色信贷, 机器学习, 风险管理

Research on Green Credit Risk Management in Commercial Banks Based on Machine Learning

Ting Wang, Li Zhang

Business School of Yangzhou University, Yangzhou Jiangsu

Received: Mar. 12th, 2025; accepted: Mar. 27th, 2025; published: Apr. 27th, 2025

Abstract

This paper explores approaches to enhance the efficiency and accuracy of green credit risk assessment by constructing a comprehensive evaluation system based on big data. A sample of 470 listed companies in China was selected, with the SelectFromModel used to screen key indicators, and the SMOTE oversampling algorithm employed to address sample imbalance issues. The research findings indicate that the machine learning model utilizing the Random Forest algorithm is more suitable for assessing green credit risk.

Keywords

Green Credit, Machine Learning, Risk Management

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

环境治理问题的不断加重让绿色金融逐渐成为金融业发展的重要方向。作为绿色金融的核心组成部分,绿色信贷使用商业银行等金融机构作为渠道为环境友好型等可持续发展项目或企业提供资金支持,目的是推动经济社会的稳定发展。绿色信贷在国际层面已经得到了广泛的应用和深入的发展。众多知名国际银行不断推出层出不穷的绿色相关信贷产品和服务,以期望通过优化信贷结构、强化风险管理等措施来积极支持环保产业的健康发展。此外,绿色信贷领域的国际合作与交流,在国际组织的大力推动下助力全球绿色金融市场的健康发展。与此同时,国内的绿色信贷也同样展现出迅猛的发展势头。随着国家对环保事业的日益重视以及“双碳”战略的稳步推进,绿色信贷已成为我国商业银行的重要业务领域之一。

绿色信贷政策的核心目标是通过引导资金流向环保和可持续发展领域,从而抑制高污染、高能耗产业的过度扩张。在这一背景下,绿色信贷信用风险评估成为商业银行在贷款决策中的关键环节。该评估要求银行依据绿色信贷政策,将环境、社会和治理(ESG)因素整合到信用风险模型中,以全面衡量企业的环境责任、社会影响及财务健康状况。然而,当前商业银行在实施绿色信贷时,普遍面临信用风险评估方法不够科学和系统化的问题。为有效管理信贷风险,亟需构建一套基于大数据的综合评估体系,并将其应用于实际业务中。这一体系不仅能够确保信贷资金流向可持续和环保项目,还能显著降低因环境和社会问题引发的潜在风险。科学的评估模型不仅有助于降低银行风险,还能提升信贷决策效率,促进绿色金融市场的稳健发展。然而,我国商业银行现阶段在绿色信贷业务的信用风险评估模型方面缺乏创新,具体体现在评估过程依赖传统经验,且用于评估风险的指标之间缺乏内在联系和相互影响,无法涵盖绿色信贷业务的所有重要方面和关键环节。因此,鉴于完善绿色信贷信用风险评估体系有着迫切的需求,着手进行创新性研究,并构建一套科学且高效的评估模型,对于推进商业银行绿色信贷业务的实际应用具有深远的意义。本文以当前绿色信贷信用风险评估领域的现状为研究起点,通过系统梳理国内外相关文献并进行深入理论分析,全面探讨我国在这一领域面临的主要问题与挑战。在此基础上,提出优化设计思路,构建评估模型并进行实证检验,旨在提升绿色信贷信用风险评估的效率和准确性,最终为我国商业银行绿色信贷业务的发展提供切实可行的政策建议。

2. 文献综述

绿色信贷信用风险评估作为风险管理中的一环,对商业银行发展绿色信贷业务以实现利润最大化起着重要作用。最早期的信用风险评估一般都是建立在定性的基础上,没有一个客观的标准。随着时间的推移,定量研究也逐渐发展起来,目前对企业信用风险的评价主要是利用数学方法,如同统计学习、概率分析等,这些方法通常基于一系列假设,例如数据的正态分布、变量之间的独立性等。如果假设在实际应用中可能不成立,会影响模型的准确性,并且统计模型通常更注重解释性,即解释变量之间的关系,而可能牺牲一定的预测准确性。但对于信用风险评估来说,预测准确性往往是关键,因此机器学习在信

用风险评价中比传统的统计学方法更具优势。

银行业的信用风险评估业务在机器学习、大数据分析以及云计算等前沿技术的快速发展下实现了从传统人工操作向智能化主导的跨越式转型。这一重大转变引发了学术界对信贷风险管理研究的广泛关注。当前, 基于大数据技术的信用评估模型主要采用涵盖了多种统计方法和机器学习算法的分类方法。

如李进[1]构建了基于随机森林算法的绿色信贷信用风险评估模型并将它与传统模型的评估结果对比, 结果显示使用随机森林算法准确率更高。Li 与 Wu [2]采纳 BP 神经网络架构, 构建了一个信用风险评估模型, 实践证明该模型具有可行性。王思宇, 陈建平[3]认为将 LightGBM 算法运用于信用风险评估模型训练速度更快, 效率更高。Vicente 和 Marqués [4]收集了 2000 年到 2013 年之间研究企业破产和信用评级的相关文章, 使用不同的数据样本和方法对决策树, 支持向量机, K 近邻等算法展开了评估。从综合表现上来看, 支持向量机算法在文章中的表现要比其他算法好得多。张迪[5]利用模糊层次分析法和 BP 神经网络等方法为商业银行构建了一套绿色信贷风险评估方案。莫志宏, 唐迪[6]用最小误判法修正 KMV 模型检验风能和光伏行业绿色信贷风险状况。Farquard 和 Sriramjee [7]采用了一种 PCA 和 SVM 模型相结合的方法, 首先采用 PCA 模型进行降维, 然后采用 SVM 模型进行分类, 通过数据验证, 其效果比逻辑回归更好。

由此可见, 目前已有许多学者尝试构建不同方法的信用评价模型, 本文在现有研究的基础上, 试图通过比较三种不同机器学习算法的预测效率和准确率来探讨其在绿色信贷信用风险评估模型中的适用性, 为商业银行绿色信贷业务的健康发展提供科学依据。

3. 研究设计

3.1. 数据处理

本文以绿色信贷作为研究主题, 并以我国上市公司为实证分析对象。依据证监会 2012 年行业分类标准, 选取了 3351 家上市公司的 2022 年财务指标、ESG 评级以及信用评级数据。为确保数据的完整性和研究结果的可靠性, 对样本进行了严格筛选, 剔除了处于停牌状态、信用评级数据不完整以及缺少 ESG(环境、社会和治理)评级数据的企业。经过筛选, 最终确定了 470 家上市公司作为研究样本, 涵盖房地产、煤炭、电力、电子等二十多个行业。本次研究所使用的数据均来源于国泰安 CSMAR 数据库。

3.1.1. 指标的选取

建立绿色信贷风险评估指标体系是确保绿色信贷项目兼具安全性与收益性的关键环节。银行等金融机构对企业的信用风险进行评级通常是运用分类模型, 对以往的公司资料进行师表, 包括违约与非违约、破产与非破产等等, 进而找出公司特性与可能存在的信用风险的关系。这一过程需要全面评估企业的多个方面, 例如财务健康状况、环境保护绩效、管理效率以及外部环境的稳定性等。为实现科学化、系统化的指标体系构建, 本文选取了能够体现企业内部财务表现和外部环境状况的相关数据。财务指标包括偿债能力, 盈利能力, 经营能力, 发展能力以及资产负债, 其中偿债能力反映了能够按时偿还债务的能力, 是评估企业风险承受能力的重要指标; 盈利能力即企业获取利润的能力, 企业价值创造的结果取决于盈利能力的大小; 经营能力反映在成本的控制以及效率的提高等方面, 是判断企业能否长期稳定发展的关键因素之一; 发展能力是企业能够持续, 稳定经营的关键, 它是反映企业未来发展前景的重要指标之一; 企业的资产负债情况由所拥有的资产情况及所承担的负债构成, 是企业规模的表现和盈利的基础, 可以用来预测未来发展情况。ESG 评级评价企业可持续发展绩效的框架, 将他们纳入评价指标, 可以用来反映企业在环境、社会、治理这三个方面的表现水平, 同时也代表了企业信贷偿还的意愿与能力。

3.1.2. 数据预处理

为提升模型预测精度并减少指标选择的主观性影响, 本文运用 Python 中的 SelectFromModel 模型进行关键指标筛选。通过特征重要性排序, 得到如表 1 所示的结果, 最终模型选取了重要性排名前 13 的指标。整合后的绿色信贷信用风险评估指标体系如表 2 所示。在数据标注方面, 本文以国泰君安数据库中信用评级 AA 及以上的企业作为非违约组, 标记为 1, 其余为违约组, 标记为 0。针对分类问题中常见的样本不均衡现象, 本文采用 SMOTE 过采样算法对数据进行平衡处理, 处理后的样本容量详见表 3。

Table 1. Indicator importance score table

表 1. 指标重要性得分表

变量	增益	变量	增益
营运资金比率	0.0580	长期资本收益率	0.0196
E 得分	0.0500	金融负债比率	0.0194
速动比率	0.0486	应收账款周转率	0.0193
总资产增长率	0.0474	有形资产比率	0.0189
流动比率	0.0464	每股净资产增长率	0.0177
流动资产比例	0.0356	资本密集度	0.0176
非流动资产比率	0.0331	长期资产适合率	0.0175
投资收益率	0.0275	股东权益周转率	0.0172
综合得分	0.0275	流动负债比率	0.0172
资产负债率	0.0271	净资产收益率	0.0171
管理费用增长率	0.0270	可持续增长率	0.0170
销售费用率	0.0264	所有者权益增长率	0.0169
流动资产周转率	0.0259	总资产周转率	0.0168
现金比率	0.0234	营业利润率	0.0167
权益乘数	0.0229	资本保值增值率	0.0164
产权比率	0.0228	利息保障倍数	0.0158
S 得分	0.0227	总资产净利润率	0.0158
G 得分	0.0226	资本积累率	0.0156
存货周转率	0.0216	营业收入增长率	0.0151
经营负债比率	0.0196	资产报酬率	0.0151

Table 2. Green credit risk assessment indicator system after screening

表 2. 筛选后绿色信贷风险评估指标体系

指标类型	指标内容
财务指标	偿债能力 流动比率、速动比率、现金比率、资产负债率、权益乘数、产权比率
	盈利能力 投资收益率、销售费用率
	经营能力 存货周转率、流动资产周转率
	发展能力 总资产增长率、管理费用增长率

续表

	资产负债	流动资产比率、营运资金比率、非流动资产比率、经营负债比率
非财务指标	ESG 综合得分	
	环境维度得分	
	社会维度得分	
	治理维度得分	

Table 3. Data distribution before and after SMOTE balancing
表 3. SMOT 平衡前后数据分布

	违约组	非违约组
平衡前	153	317
平衡后	317	317

3.2. 实证过程

在 Python 环境下，为评估各类机器学习模型对绿色信贷风险预测的适用性，研究首先对数据进行了预处理，随后通过系统化的参数调优以提升模型性能。在此基础上，选取了三种机器学习模型进行训练，并通过对比其分类效果，分析它们在绿色信贷风险评估任务中的适用性与性能差异。

3.2.1. 基于随机森林的绿色信贷违约风险预测

随机森林是一种集成学习算法，其核心思想是通过整合多棵决策树的预测结果来提升模型的整体精度和稳定性。在随机森林模型的构建过程中，每棵决策树的分支生成首先从原始特征集中随机抽取一个子集，该子集的规模通常小于总特征数，并可通过参数进行调整。随后，基于特定准则(例如信息增益或方差最小化)从子集中筛选出最优特征用于节点划分。这种基于最优特征的分支策略，使得决策树能够更有效地对样本进行分类或更精确地拟合目标变量的变化趋势。

3.2.2. 支持向量机的绿色信贷违约风险预测

支持向量机(SVM)是一种基于统计学习理论的分类与回归分析方法。其基本原理在于识别数据集中的关键数据点——支持向量，这些数据点位于数据分布的边界附近。通过这些支持向量，SVM 能够构建一个超平面，该超平面在模型中具有核心作用，作为划分不同类别数据的决策边界。为了实现这一目标，SVM 通过不断调整参数来优化超平面的位置，使得不同类别的样本之间的间隔最大化。

3.2.3. 基于 XGboost 的绿色信贷违约风险预测

XGBoost 算法的基础是梯度提升算法，它通过迭代地添加新的基学习器(如决策树)来最小化损失函数，从而提升模型的预测精度。每个新添加的基学习器都是为了拟合前一轮模型的残差而设计的。残差是指模型在当前迭代中的预测值与真实值之间的差异。通过不断迭代，每一轮都试图通过新的基学习器来减少这个残差，从而使得整个模型的预测结果更加接近真实值。

3.3. 结果分析

在参数设定方面，随机森林算法的关键参数涵盖分类器数目及其深度。一般而言，分类器数量的增加会提升准确率，但增至某一临界点后将趋于平稳。通过测试 10、40、50、70、100 及 200 个分类器的配置，我们发现当分类器数量达到 50 时，准确率趋于稳定。随后，我们进一步调整了最大深度参数，经过多次试验，确定当随机森林配置 50 个分类器且最大深度设为 25 时，模型的分类性能最佳。至于支持

向量机, 我们选择了线性核函数, 并将正则化参数设定为 0.1。对于 XGBoost 算法, 当基础分类器数量为 50 且最大深度为 2 时, 其预测效果最为理想。三种模型的 AUC 曲线如图 1 所示, 根据所得预测结果, 通过 ROC 曲线评估模型整体性能时, XGBoost 与随机森林的曲线下面积基本相当, 这一结果从侧面验证了 XGBoost 与随机森林的表现最为优异, 模型的整体效能达到了最优水平。

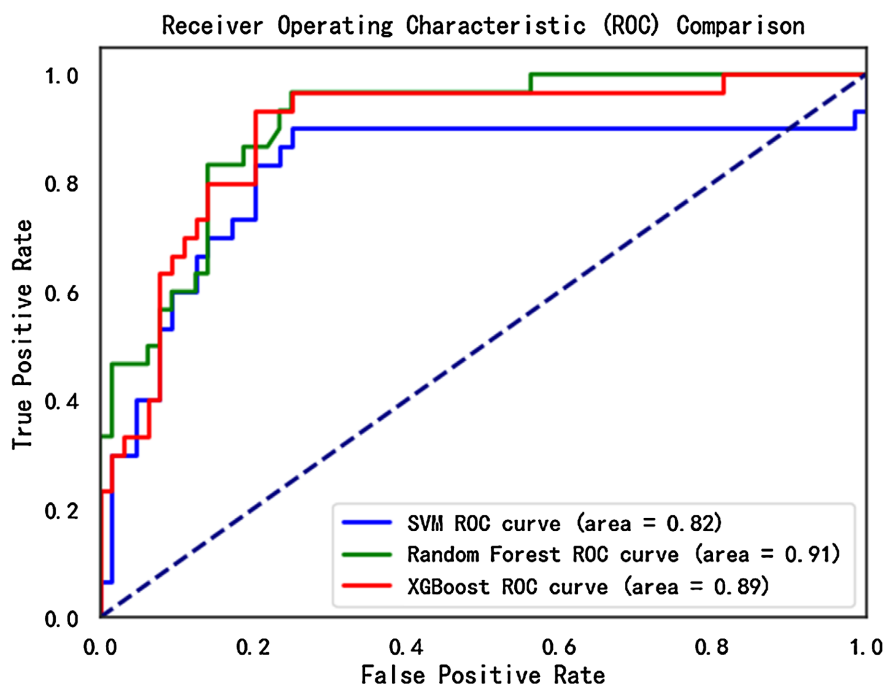


Figure 1. AUC curve of three models
图 1. 三种模型 AUC 曲线

在银行实施绿色信贷政策并进行企业信用风险评估的过程中, 其首要目标是精确预测企业未来的违约风险。为达成此目标, 银行必须确保所使用的评估模型具有较高的准确率, 从而能够可靠地判断企业的违约可能性。对于那些被预测为存在违约风险的企业, 银行应谨慎对待其贷款申请, 以减少潜在的信贷风险。同时, 针对信用评估表现优异的企业, 银行应当加大信贷支持力度, 此举不仅有利于推动企业的长远发展, 还能进一步提升银行的盈利效能。另外, 银行还需特别着重于精准识别违约风险较高的企业, 以保障资金的安全性, 有效遏制不良贷款的形成。因此, 在各项指标中, 准确率最为重要, 综合对比各学习器预测效果可知, 在准确率上, 随机森林违约组预测准确率为 0.74, 在所有学习器中排名最高, XGBoost 次之; 此外, 随机森林在所有指标上都表现最好, 尤其是在违约组的准确率和 F1 分数上。这意味着随机森林在处理这个特定的预测任务时, 能够更好地捕捉数据中的复杂模式。所有模型在非违约组的准确率和召回率都高于违约组, 这可能是因为非违约组的样本数量较多, 导致模型更容易学习到非违约的特征。召回率和 F1 分数对于信用风险预测尤为重要, 因为它们直接关系到模型识别违约案例的能力。由此可见, 随机森林模型在绿色信贷风险评估上有很大的作用空间, 在绿色信贷风险评估中具有更高的适用性, 能够较好地应用于风险评估预警中。

4. 结论

在当前绿色信贷信用风险评估框架内, 部分模型由于显著受到主观判断的影响, 其评估结果的客观性与可信度备受挑战。与此同时, 另一些模型则面临过拟合问题, 这一局限阻碍了它们在多种情境下的

Table 4. Comparison of prediction results across models
表 4. 各模型预测结果比较

算法	组类	准确率	召回率	F1 分数
随机森林	违约组	0.74	0.67	0.70
	非违约组	0.85	0.89	0.87
支持向量机	违约组	0.71	0.67	0.69
	非违约组	0.85	0.88	0.86
XGBoost	违约组	0.72	0.60	0.65
	非违约组	0.83	0.89	0.86

广泛适用性及泛化潜能的发挥。对传统评估指标体系进行深入剖析时，我们发现指标选取过程中普遍存在显著的主观倾向，并且可能存在未能全面覆盖所有核心风险要素的情况。这种局限性可能削弱评估结果的全面性与精确度。本文首先将非财务指标纳入评估体系，作为全面衡量企业绿色绩效的一个维度。通过实证分析验证，相较于传统体系，可以更深入地反映企业在环境与社会 responsibility 方面的表现。其次，本研究开创性地将机器学习技术融入绿色信贷领域，通过对比分析三种不同算法的测试结果(如表 4)，最终鉴定出由随机森林算法构建的绿色信贷信用风险评估模型，展现出了卓越的评估效能，并具备高度的实践应用潜力。在针对上市公司绿色信贷信用风险的实证分析中，该模型与支持向量机(SVM)及 XGBoost 等传统评估模型相比在削弱主观因素干扰、增强评估精确度方面表现尤为突出，其总体准确率位居所有模型之首，这进一步有力验证了该模型的可行性与有效性。

基金项目

扬州大学商学院研究生创新项目，项目批准号：SXYJSCX202437。

参考文献

- [1] 李进. 基于随机森林算法的绿色信贷信用风险评估研究[J]. 金融理论与实践, 2015(11): 14-18.
- [2] Li, G. and Wu, Y. (2010) Empirical Research about Credit Risk on Neural Network Based Bp Algorithm. 2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering, Kunming, 26-28 November 2010, 461-463. <https://doi.org/10.1109/iciim.2010.431>
- [3] 王思宇, 陈建平. 基于 LightGBM 算法的信用风险评估模型研究[J]. 软件导刊, 2019, 18(10): 19-22.
- [4] García, V., Marqués, A.I. and Sánchez, J.S. (2014) An Insight into the Experimental Design for Credit Risk and Corporate Bankruptcy Prediction Systems. *Journal of Intelligent Information Systems*, **44**, 159-189. <https://doi.org/10.1007/s10844-014-0333-4>
- [5] 张迪. 我国商业银行绿色信贷风险评估研究-基于模糊层次分析法和 BP 神经网络算法[J]. 山西农经, 2019(5): 147-148.
- [6] 莫志宏, 唐迪, 葛林. 基于 KMV 模型绿色信贷风险管理研究-来自风电和光伏产业的经验数据检验[J]. 管理现代化, 2015, 35(1): 97-99.
- [7] Farquard, M.A.H., Ravi, V., Sriramjee, and Praveen, G. (2011) Credit Scoring Using PCA-SVM Hybrid Model. In: Das, V.V., Stephen, J. and Chaba, Y., Eds., *Computer Networks and Information Technologies*, Springer, 249-253. https://doi.org/10.1007/978-3-642-19542-6_40